DOCUMENT RESUME

ED 166 251 TM 008 291

AUTHOR Kennedy, Dale
TITLE The Nebraska Blueprint for School Assessment Programs. Volume 2: Processes and Procedures.
INSTITUTION Nebraska State Dept. of Education, Lincoln.
PUB DATE 78
NOTE 56p.; For related documents, see TM 008 290 and 292

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Data Processing; *Educational Assessment; *Educational Testing; Elementary Secondary Education; Guidelines; *Program Administration; *Recordkeeping; Scoring; *Student Records; *Testing Programs; Test Interpretation

ABSTRACT

This is the second of three volumes designed to cover the problems and issues of developing and managing a school assessment program. The series offers guidelines and illustrations of several special aspects of assessment programs, such as administering tests and documenting their results. This volume deals with the planning of assessment programs; the program's standing documentation; and the problems of administering, scoring, record-keeping, and reporting--referred to collectively as the logistics of management. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166251

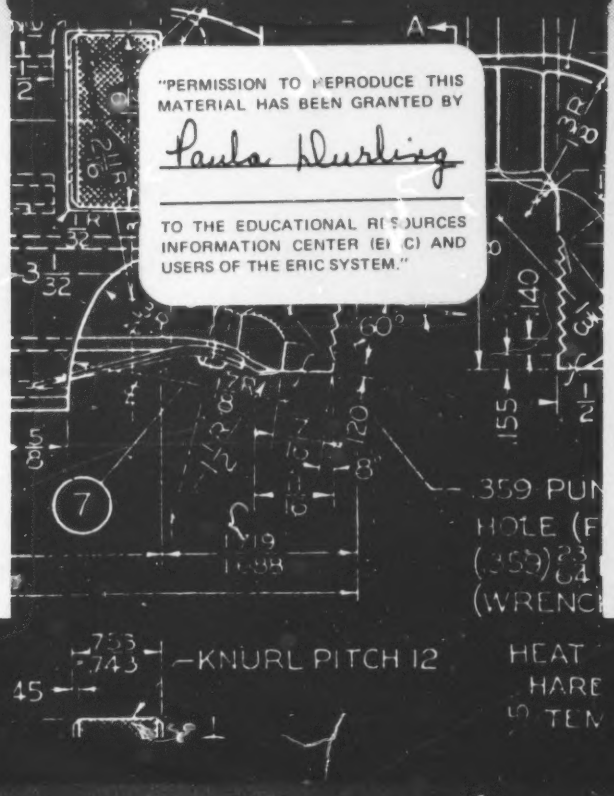
U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Paula Hurling

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."



The Nebraska

BLUE PRINT

For School Assessment Programs

Volume 2: Processes and Procedures

Issued by the Nebraska State Department of Education

M. Anne Campbell, Commissioner of Education

301 Centennial Mall South
P. O. Box 94987
Lincoln, Nebraska 68509

1978

TABLE OF CONTENTS

Volume II: Processes and Procedures

1. Introduction 1
   - The Planning Process 1

2. Design 3
   - The Administrations Chart 4
     - Rules of Thumb 4
   - The Expansion of Chart Entries 6
     1. A Form for Evaluating Tests or Assessment Techniques 8
     2. Procedures 10
     3. Specific Uses 10
     4. Justification 12
     5. The Long Range Value of the Design Statement 12

3. Logistics of Management 19
   - A. Administering 20
   - B. Scheduling 20
   - C. Environment 21
   - D. Staff Support 21
   - E. Training Examiners 22
   - F. Distribution of Materials 23
   - G. Quality Control 23
   - H. Examiner Checklists 24
   - I. Scoring 27
     - Types of Scores 28
       1. Raw Scores 28
       2. Expectancy Tables 29
       3. Percentiles 29
       4. Standard Scores 30
       5. Normalized Test Scores 30
       6. Criterion Scores 30
       7. Error of Measurement 30
       8. Profiles of Part Scores 31
       9. Item Analysis 32
   - J. Record Keeping 33
     - Automated Record Keeping 33
     - Explanatory Text 34
     - Long Term Storage 35
   - K. Reporting 36
     - Common Points of Information 37
     - Reporting Formats 37
       1. Cumulative Record File—Individual Reports 37
       2. Group Summaries 38
          - a. Class Lists 38
          - b. Trend Tables 39

c. Comparison Tables 39
d. Distributions 39
e. Performance Sequences 40
f. Composite Subject Charts 40
4. Elements of Interpretation 41

COMMISSIONER'S FOREWORD

Assessment, in ordinary words, is the effort devoted to finding out if students are learning what we think we are teaching. Like instruction, it is a complex undertaking, both in conception and in practice. Both aspects, instruction and assessment, are essential to the process of education and both require careful and thorough development to be of best effect. The Department's Blueprint for School Assessment Programs covers the problems and issues of developing and managing an assessment program. It also offers guidelines and illustrations of several of the more special aspects of assessment programs, such as administering tests or documenting their results. The Blueprint's coverage is comprehensive and will have value for a variety of audiences that may be involved in assessment programming.

It is anticipated that the Blueprint will prove a useful reference in planning, developing, evaluating, reviewing, and managing local assessment programs.

M. Anne Campbell
Commissioner of Education

AUTHOR'S FOREWORD

The Blueprint is intended to be a repository of theoretical and practical information on student assessment and the problems of assessment programming. It has the qualities each of a textbook, a handbook, and a reference manual. It is possible for it to be read and used at several different levels. It could satisfy readers having general interest in assessment theory and technology, but it should also prove to be of repeated value for readers who are actually involved in assessment programming, such as the planning committee members, examiners, and the persons charged with the school or district's assessment responsibilities. It may be scanned for general information or it may be studied for precise and technical information. It can serve to generally refresh and improve one's overall understanding of the issues and problems of assessment, or it can be employed as a guidelines and procedures manual for the actual planning and operation of an assessment program.

Those persons who are actually involved in the planning and operation of a school or district assessment should seriously consider acquiring and reading Standards for Educational and Psychological Tests, published by the American Psychological Association, Guide For School Testing Programs, published by the National Council on Measurement in Education, and Essentials of Psychological Testing, by Lee J. Cronbach, published by Harper and Row. There are other good books on psychological testing besides Cronbach's. The Essentials, however, exhibits the unique combination of well explained theory along with thorough discussions of practical concerns. Cronbach also writes in a scientific style

that is both exacting and conservative in statement. These are qualities which the Blueprint also attempts to achieve, though in a more synoptic form. The Blueprint, together with these three references, should soundly provide a thorough and comprehensive knowledge of assessment theory and practice.

Dale Kennedy

PREFACE

Human children are growing, learning, maturing, thriving organisms. The ingredients of their perceptions and cognitions, of their learning, far exceed those of other living things found upon earth, for children's growth must take place on several levels simultaneously: physical, intellectual and emotional. The purpose of education, then, is to shape and coax this growth (if not to stimulate it outright). Education, when harnessed to growth, leads it to its most mature levels in a way that ideally enhances further self education. An education "completed" ideally leaves the student open for a multitude of future circumstances.

Learning takes place as a function of three things: 1) past learning, in which one can include native ability because its effects are in fact manifested by the past learning; 2) the occasion upon which the learning takes place, i.e., the learning opportunity to the teacher and the learning experience to the student; and 3) the consequences contingent upon the learning behavior itself. Learning occurs if the behavior of the students takes a new shape during and after the occasion.

The change is lasting or strong, occurring again on appropriate future occasions or affecting future learning, in proportion to the reinforcement of the learning on the occasion of first change. But the student must first perceive the change and the consequences thereof. If the consequences are natural, as in the discovery of the world of literature through reading, the student must value these new possibilities. Teachers must also perceive the change if their actions toward the students are to be consistent, timely, and in pace with the students'

rate of growth. This reading of student behavior and ability is assessment.

Most assessment is informal, taking place each day, any time the teacher and student interact. Most of the student's self-assessment also takes place on informal occasions. The assessment function becomes formal when the student is asked to perform on occasions which would not normally be encountered as a daily routine, designed to reveal something of the student's ability. These occasions of formal assessment are the subject matter of the Blueprint.

The Blueprint is published in three separate volumes. Volume I, Purposes and Categories of Measurement, deals with the regulations governing school assessment programs, the general purposes for which assessment information is collected, and the kinds of measuring instruments used for such collection. This is the most conceptually involved volume, but it does provide a comprehensive view of existing measurement theory and measurement technology with respect to educational settings.

Volume II, Processes and Procedures, deals with planning of assessment programs, the program's standing documentation, and problems of administering, scoring, record-keeping and reporting, referred to collectively in the Blueprint as the logistics of management. The information provided in this volume balances the conceptual content of Volume I with detailed exposition on practical considerations. Volumes I and II are intended to complement each other.

Volume III, Technical Information, is nothing more than a repository of useful information of a technical or abbreviated nature. It has sections concerning reliability and validity, addresses of test publishers,

lists of references, computing guides for various statistics, test tables and a glossary. The information in this volume is also of a practical nature, even more so than Volume II.

There are four central assumptions underlying all volumes of the Blueprint.

First, a district's assessment program already exists and stands in a working relationship with the operations of instructional delivery. The primary object of the Blueprint is to provide the concepts and procedures that will bring the documentation of the program to a comprehensive, organized form, if it does not already exist in such a systematic fashion.

Second, the assessment of individuals and groups must be fair to those individuals and groups. If a test discriminates against an individual on the basis of background or physical handicap, the results will not represent true information on the individual's abilities or achievement. If used to make decisions about the individual's future education, such unfairly obtained data would be misinformation causing mistakes in judgment that may result in the miseducation of the individual. The fairness stricture also applies to mainstream groups. Here, if the administrations are poorly controlled, the margin of error is enlarged and the test becomes an inaccurate measure mostly containing misinformation. Assessments must be exactly administered if they are to be fair. Miseducation is the consequence of mistaken judgment.

The possibilities of miseducation are indeed some of the central reasons for establishing and maintaining an assessment program. For example, it is always desirable to fit instruction to known strengths

and weaknesses of the student, or groups of students. The educative process would be expected to reinforce the strengths, improve the weaknesses, and in its combination of effects stimulate an optimum amount of learning. But just as surely as instruction can promote optimum growth, it can drop a ceiling on learning. Instruction can mismatch, as well as match, the strengths and weaknesses of students. And in so doing, it fails to promote growth where potential is great or expects more of the student than the student is truly capable of. A mismatch guarantees a growth far less than what is optimum and, in effect, blocks such growth. Miseducation is simply the diminishment of learning through an inappropriately applied (indeed, an alleged) educative process. It is always a danger. Good assessment information should reduce and limit occurrences of this plight.

Third, grade promotion and graduation sanctions, contingent upon assessment information, are not necessary to effective and useful assessment programs. They are essentially a special credit system extended beyond the usual credits for attendance and passing courses. Where they are instituted, they became a defining feature of a whole assessment program, yet a program has several good uses without them. The logic and rationale of grade promotion and graduation requirements are not topics of the Blueprint. The subjects that are discussed, however, would give a soundness to the assessment information, so that if used for promotion, the information would at least be validly obtained.

Fourth, the readers of this manual and all those involved in assessment, either as tester, subject, or interpreter, should understand and appreciate the scientific nature of educational assessment. An interpre-

tation made from assessment information is a work of scientific reasoning. Both the proponents and the critics of scientific method as applied to education tend to regard the mechanical qualities of the methods of science as its only virtue, or weakness, as the case may be. This neglects the deliberateness and natural conservatism of scientific method. It also misses the cultivation of balanced perspective. Whether very mechanically obtained, or more informally so, assessment information must: first, be the right information needed; second, be accurate; and third, be viewed in relation to other assessment information on the same individual or group. If the information does not fit the questions asked, or if it is error-ridden, then it is misinformation and, if interpreted, would foster mistaken judgments and lead to the probable miseducation of the students. It requires an understanding of scientific method to recognize when and under what conditions the information gathered for the purposes of assessment proves to be misinformation, data which can be harmful to the future educational prospects of students.

Above all else, this scientific perspective must be grounded in a concrete understanding of the conditions under which the data are actually produced. It is the children who face the test and perform as best they can that day. The temperature of the room, the lighting, their breakfast or lack of breakfast, and many other daily, mundane facts, all conspire to affect the children's performance. The mechanical rigors of administration are intended to even out the disadvantages and to heighten the individual's performance. A testing procedure cannot be too precise when these are its objects.

INTRODUCTION

The Planning Process

Whether an assessment program is beginning, being redeveloped, or simply maintained, it is still fundamentally the result of a planning process. The assessment committee charged with the planning responsibility should represent teachers, parents, administrators and counseling or testing personnel. Lengthy tenure on the planning committee is desirable, as this best assures a stable assessment program. Changes are inevitable, but if the committee has considerable experience in making assessment decisions, the changes will more often be made with deliberation, and with greater understanding of the standing problems of assessment. Further, the committee will be best able to interpret significant trends in test results if its members possess knowledge of the past choices, reasons, and decisions about the nature and intent of the assessment program. Finally, the benefits of in-service training on assessment will be concentrated when committee members remain in its service for a goodly length of time.

There are two bodies of knowledge which the committee should develop, maintain and review. First is knowledge of the curriculum and the school's goals for students, couched in whatever concepts and language employed by those responsible for the articulation of the curriculum, e.g., "learning outcomes," " objectives," "competencies," etc. Although an assessment committee typically will not produce documentation on the curriculum, this information is vital to any rational approach to the design of an assessment program. Assessment information produced from a

design that poorly matches the intentions of the school is non-information, or worse, misinformation, setting up mistaken judgments.

The second body of knowledge concerns all the assessment activity presently conducted by the school. This knowledge has three aspects. First is the design of the program which details the instruments used, to what purpose, for whom, and when. The second aspect is the logistics of management involving the procedures for administering the instruments, scoring the results, keeping the records, and making reports of the results. The third aspect is the interpretation and use which focuses upon the audiences which will receive reports of the results--the students, their parents, administrators, teachers, counselors and the public. Each of these three aspects can be developed into written policy and procedure. Such documentation would also serve to justify the committee's choices in planning the assessment program, as well as to provide a clear description of exactly what the program consists. Critics would then have to become as well informed as the planners to be creditable.

The remainder of the Volume II is devoted to developing each of these aspects of assessment: the design; the management logistics; and the interpretation and use of the results. The discussions will present both the central concepts and problems of each aspect and the guidelines and procedures that can be used to plan, manage, and document the assessment program overall.

DESIGN

Given the familiarity with the general purposes of assessment, the varieties of tests and assessment techniques available and their particular uses, any group committed to the planning process will desire to have a clearly defined product which finalizes its decisions and presents them in an ordered fashion. This would be the design of the assessment program. The design statement will have two written parts. The first is the Administrations Chart, detailing which instruments are to be administered, to whom they will be administered, to what purpose and when. Since the use of such a chart would extend through many years, it is more convenient to present this feature in six parts: preschool; lower elementary; upper elementary; intermediate; secondary; and adult.

Each entry onto the purpose-by-level charts will be further expanded by the evaluation form, in full, of the specific tests or techniques, as well as by a statement defining logistical procedures and specific uses and the ends to which the results are to be put. This expansion of detail is the second part of the design statement and should also include a statement justifying the choice of tests and the specific uses intended.

Together, the two features present a comprehensive document describing and explaining the assessment program. It would in fact represent, in particular, the blueprint for the local assessment program. These two parts of the design statement are illustrated in the following pages. The Administrations Charts are fully illustrated, but the entry expansions are abbreviated because of length. A more detailed discussion of the procedures for scheduling, administering, scoring, record-keeping,

and reporting that could be presented in the Chart Entry expansion is provided in the management logistics sections.

The Administrations Chart

The chart itself can be presented in a variety of particular graphic layouts. It should in all cases, however, divide the column and row headings into the divisions of grade and the types of purposes, respectively. The entries within the body of the chart specify a particular test and will be positioned according to its purpose and the grades to which it will be administered. The entries provided by the illustration below, however, are not the names of specific tests, but rather generic types of tests.

Administrations Charts can summarize the assessment for either the particular student body or the district overall. Its exact nature depends on the local district's situation. There are a few general guidelines or rules of thumb that can be helpful in making some of the choices for chart entries.

Rule 1

Restrict norm-referenced instruments to diagnostic purposes wherever possible. Inferences about learning progress using norm-referenced scales are difficult pieces of reasoning. Judgments of strength and weakness, however, can be soundly and simply made from the comparison of an individual's performance to the group averages.

Rule 2

Anchor as many progress instruments to criteria or performance standards as possible. Interpreting change in norm-referenced score is confounded by a number of difficulties not shared by a criterion-referenced score. First, the distances between points on a norm-referenced scale are known to be unequal and, consequently, the quantity of change will differ between intuitively equal intervals. The real change in a gain from the twentieth to the fortieth percentile is greater than a gain from the fortieth to sixtieth percentile. This caution is not applied to criterion-referenced scales.

Rule 3

Use short, single composite scores early in the student's career and move to multiscore profiles at later ages. Attention, concentration, and patience are less developed in younger children. The longer they endure testing conditions, the more irrelevant motivational factors will influence the results. Multiscore profiles are generally longer and require several kinds of response sets (type of items) to complete. Early ability is already known to have low correlations with future performance because of the spurts and upheavals of early intellectual maturation. The overtaxing of the child's abilities with lengthy, well-factored batteries in one sitting adds to the imperfection and further threatens the validity of the interpretation.

Profiles can instead by built up through a dispersed sequence of tests and assessment tasks over a several month period. For example,

series of short criterion-referenced tests or a series of applied performance tasks could be given over a three-month span.

Rule 4

Place the diagnostic assessment of a given kind well before the progress assessment of the same kind, e.g., a basic skills diagnostic battery in the fall before a basic skills achievement battery in the spring.

Rule 5

Stretch the intervals between like kinds of testing. The effects of instruction, even when well-grounded on diagnostic information, are gradual. Growth does not have to be checked constantly to be accurately measured. The growth in learning will be more pronounced the longer the interval between testings. A three year cycle, for instance, would provide a large measure of instructional intervention. Yet, by the end of students' career, they have been thoroughly measured on growth and progress four times, frequently enough to make any trend visible.

The Expansion of Chart Entries

Each chart entry is a particular test or assessment technique which will have a host of particular characteristics that affect the way the test is used. It will also require certain kinds of administering procedures necessary to its valid application. The expansion of the chart entries to include this information can be carried out in more or less

Illustration 1 - Administrations Chart,
Lower Elementary Grade

	Grades					
General Purpose	1		2		3	
	Fall	Spring	Fall	Spring	Fall	Spring
Progress		Reading test: Vocabulary, Simple Comprehension				Lower Level Basic Skill Battery, Criterion Referenced
Diagnosis	Ready Readiness Test		Lower Level Basic Skills Battery, Criterion Referenced		Intelligence Test	

Learning Disabilities Procedur

detail. It should, however, be sufficient to explain the nature of the test, its critical aspects of administration, its expected uses, and justify its selection over alternatives.

The illustrations of such expansion provided in the following sections are intended to show the range of the kinds of information that can be catalogued. Which kinds of information and in how much detail it will be given are questions that must be answered locally. Expansion should be sufficient to satisfy the local community demand for public information and yet be thorough enough to well inform other professionals in the field.

1. A Form for Evaluating Tests or Assessment Techniques

Whether to choose tests or techniques, or whether to justify and document existing choices, the planning group will need to carefully evaluate each test. An item-by-item form covering each of the following points will organize and order the varied information available on each test or technique. This organized information aids in the selection process and will document those choices made. The planning group will need to obtain copies of the test manuals and specimen sets as well as other related information concerning the test.

Points of Information

Title

Author

Publisher

Date of publication

Date of norms, if any

Number of forms

Groups for which assessment is applicable

General purpose of the testing

Specific uses

Categorical type of test or technique (the scheme of categories outlined in the "Varieties of Tests and Assessment Techniques" could be used)

Mental qualities or behavioral capabilities represented by scores

Number of items

Time required to take the test

Description of:

i. The test

ii. Author's purpose

iii. Author's basis for selecting items

iv. Items (samples)

v. Scoring

vi. Report format

Reliability - type of coefficient and description of conditions under which it was obtained

Validity - applicable types; if criterion-related, includes coefficients and description of conditions under which they were obtained; if content-oriented, describes method of local validation

Ease of administration and scoring

Adequacy of directions

Training required for administering

Costs

References and reviews (Buros' Encyclopedia of Mental Measurement is an excellent start)

2. Procedures

The actual management of the assessment program may be the responsibility of the entire planning committee, but usually it will fall to the member who is charged by the district for overall coordination. The person may be a test director, a school psychologist, a counselor or an administrator. This management responsibility will essentially involve the execution of procedures that are somewhat mechanical in nature, e.g., getting the testing materials to the right persons on the right dates and getting them back again, and so on. They are aptly considered the logistics of the assessment program.

The logistical aspects of assessment are discussed in detail later. The value in documenting the critical aspects lies in the eventual need to transfer the responsibility for them to new individuals. When the logistics are a matter of record, the new individual is saved the labor of working them out for himself for the first time. Further, if the planning committee periodically reviews and upgrades the procedures, the new individual, in consulting the written document, has the benefit of a cumulative wisdom.

3. Specific Uses

A specific use of a test is defined by the audience receiving the results and by the types of decisions or judgments for which that audience can use the information. Each audience will have different levels

of understanding and different needs for the information. It is important that the reporting formats match the audience's level of understanding and needs.

The explanatory material accompanying the results should also reflect these considerations, but it must still be sufficient to ground a valid interpretation (see "Elements of Interpretation"). A more detailed treatment of test use is available in the Standards for Educational and Psychological Tests, "Standards for the Use of Tests," pages 56-73.

There are six principal audiences which may receive, view and use test results: students; counselors; teachers; administrators; parents; and the public. The types of decisions and judgments each audience may make, given the information, are listed below.

Individuals - self appraisal; educational and career planning.

Counselors - identifying individual student needs; counseling individuals in educational and career planning.

Teachers - self-evaluation; identifying individual and group needs; instructional planning; monitoring student achievement.

Administrators - identifying special group needs; selection and classification of students; instructional and curriculum (program) evaluation; recording and documenting student achievement.

Parents - counseling and home support; confidence in school; knowledge of children's achievement.

Public - school accountability; evidence of student body progress and achievement; recognition of special district needs.

Given the importance of the specific uses of assessment information,

it can be quite helpful to the planning process to begin by detailing the expected and desired uses, interviewing members of each audience, and working back from this information to the choice of the assessment instruments.

4. Justification

Each feature of the design of an assessment program is open to the questions of "Why?" and "To what purpose?" For example, why administer this particular test instead of others? Why is the information needed or useful? What value does the information have to this group or that group? Some of the questions are answered when the specific uses are detailed, but many are left open. Since they invariably are valid questions, it is useful to anticipate them in writing an assessment design.

A justification statement need only consider as many points of recommendation and explanation as the planning committee deems necessary. Attempting to anticipate all possible lines of criticism would prove burdensome. In addition to justifying decisions and choices, the justification statement also serves to bring all the key points in favor of a test, both general and specific, together in a few encapsulating paragraphs.

5. The Long Range Value of the Design Statement

The design statement is a standing document. If it includes a complete set of Administrations Charts and carefully expanded chart entries, it is also the complete catalogued record of the assessment

program--a blueprint, in effect. The charts and entry expansions should be kept as separate physical entities and be stored in looseleaf form. Individual sections then may be revised, replaced, or deleted in separate deliberations without necessitating a revision and retyping of the entire document.

First decisions, though they may suffice for the moment, are frequently improved as experience is gained. As the planning committee gains experience, as its understanding of the assessment problems grows and as it observes the results of its first decisions, better choices and wiser procedures will suggest themselves. As these changes are made and documented, the design statement is correspondingly modified. Over time, then, if the design statement is so upgraded, it stands for the best decisions and ideas that have evolved during the history of the local assessment program. Every educator fears the spectre of the reinvented wheel. The best assurance that this will not happen is to keep a written history and to see that it is read and known by those responsible. The design statement can easily be kept as an historical record, and if maintained systematically for a period of five years or more, it will serve as a basis of perspective and insight to those making the decisions affecting the nature and shape of the assessment program.

Illustration 2 - Chart Entry Expansion, Sample
Evaluation Form

Illinois Test of Psycholinguistics Abilities (ITPA), Revised Edition

Samual A. Kirk, James McCarthy, Winnifred Kirk.

The University of Illinois Press.

Revised edition in 1968.

Specific Purpose

To identify specific communication abilities and disabilities in order to delineate appropriate remedial and instructional programs.

Age Levels

Two to ten years.

Parts

12 subtests.

Description

Testing is set up for these 3 dimensions of cognitive abilities:

1) Channels of Communication
2) Psycholinguistic Process
3) Levels of Organization

Functions are tested in these 2 major areas:

1) Representational Level
2) Automatic Level

Sample Items

Sentence completion of an incomplete analogous statement, such as: "I cut with a saw, I pound with a _____." Child is instructed to complete 2nd statement appropriately.

Scoring

A composite Psycholinguistic Age (Composite PLA) is derived from a Composite Raw Score (sum of 10 basic subtest scores). This is a global score similar to a M.A. (mental age) and is considered by the authors of the test as an overall index of psycholinguistic abilities. The Composite PLA can be divided by the child's C.A. (chronological age) and multiplied by 100 to derive Psycholinguistic Quotient (PLQ) comparable in nature to a ratio IQ.

Scaled Scores (SSs) are transformations of Raw Scores such that at each age and for each of the 12 subtests the mean performance of the standardization group was equal to a SS of 36 with a standard deviation of 6. PLA norms were derived from only group means, but SSs take into account both group means and variance. SSs provide a more versatile tool for comparing a child's performance from subtest to subtest than do the PLA's.

***Author's suggestion: A SS subtest ten points below the individual's mean SS indicates an educationally significant disability. SSs from 7-9 points below the individual's mean indicate borderline disabilities.

Time of Administration

One hour.

Equivalent Forms

None

Norms

Norms were derived from the responses of approximately 1,000 children between ages of 2 and 10. This sample was selected to reflect average performance on traditional measures of intelligence, school achievement, and socio-economic status, among children with normal motor development and sensory performance.

Reliability

Retest of ITPA's reliabilities for the 12 subtests range from .12 to .86 with a median of .50. The corrected-for-range estimates range from .28 to .90 with a median of .71. The retest correlations for the composite score are .83, .70, and .70, respectively, for the three age groups.

Illustration 3 - Chart Entry Expansion,
Procedures Form

Scheduled Administration

9:30 - 10:45 a.m., October 13th, to all third grade students.

Administering

Tests to be distributed October 11th, with an administrative memo, to the teachers of all third grade classrooms. Teachers administering the test for the first time should be in-serviced at least one week in advance on the techniques required for valid administration. All test and answer sheets will be collected by a school administrator at 11:00 a.m.

Scoring

Answer sheets sent to publisher's scoring service.

Recordkeeping

Individual profiles of subtest and composite scores returned are entered on a 8 1/2" x 11" page and inserted into the individual student's cumulative folder. A copy is also filed in the principal's office.

Reporting

The students and parents are notified that the test results are in and are encouraged to schedule a conference with the school counselor to discuss the results.

Teachers are expected to review the individual results stored in the principal's office for each student in their classroom.

Group summaries are tabulated in the principal's office and a copy is to be sent to the central district offices. Copies of the group summaries will also be made available to the teachers.

Illustration 4 - Chart Entry Expansion, Sample
Specific Uses Table

Audience	Use	
Student	1.	Self-appraisal of linguistic capabilities
Parent	1.	Informed judgment as to child's linguistic capabilities
Teacher	1.	Identify strengths and weaknesses of individual students
	2.	Identify strengths and weaknesses of class performance
Administrator	1.	Identify strengths and weaknesses of class performance
Counselor	1.	Identification of students potentially needing special education
Public		Not applicable

Illustration 5 - Chart Entry Expansion, Sample
Justification Statement

The Illinois Tests of Psycholinguistic Abilities were chosen to provide diagnostic information about the language abilities of students before they enter the upper elementary grades. The division of the abilities into 12 subtests was considered sufficient and ample for making informed decisions without over-reliance on a single score. The psychological interpretation of the subtests is considered to be understandable by most persons. Particular features recommending these tests over others were: short administration time; high reliability estimates; and the fact that the tests were designed to assess disability as well as ability.

LOGISTICS OF MANAGEMENT

Once the decisions have been made as to what to measure with which test, the entire bent of the assessment effort is the control of those factors affecting the individual's performance on the test and those factors affecting the eventual use of the results. In the first case, "control" means to insure that conditions under which the test is administered are fair and well fit the test manual's instructions. This would assure that it is valid to interpret the results. More often than not, it is better to have no information than to have information invalidly obtained. In the second case, "control" means to process the raw data quickly and in an understandable form, and to deliver it to those persons who could make use of the information. This is as much a question of timing as of understanding. Persons receiving the test results certainly must understand the purposes for which the information was obtained, the properties measured, the choice of administering techniques, and the type of scores reported in order to appreciate the value of the test results. Yet, when the information can have value in the formation of instructional strategy, both for groups and individuals, it is important that the information be available before the hour of decision. It is especially disheartening to learn of information that would have made a difference in a decision already made and acted upon.

The assurance desired in both these cases is had by attention to the practical considerations of procedure and communication, of what to do and how to do it--logistical matters. Their careful management sets the stage for interpretation and use. These matters will be considered in five parts: administering; scheduling; scoring; recordkeeping; and

reporting.

A. Administering

The details of the administration of tests or assessment techniques center on one end--the control of the factors affecting the individual's performance on the test. The conditions under which the test is taken should, one, elicit the individual's best performance under the circumstances, and two, be fair and equal, each individual experiencing the same degree of advantage or disadvantage. Neither case can tolerate too great a margin of error or deviation before the results become uninterpretable and of no value, except as misinformation. The best intentions and most enlightened uses come to naught, or abuse, when the instruments of assessment are improperly administered.

The decisions and plans for administering should be set several weeks in advance of the coming testing cycle.[1]

B. Scheduling

The examination dates and times for each test or assessment technique should be set for each school several months in advance. These decisions should follow consultation with the school administrators in order to avoid special athletic events, field trips, fire drills, parent days, holidays, and other happenings that would dominate a day's school activities. Once set, the administrators and the teachers involved should receive periodic advance reminders of the coming test administra-

---

[1]What follows is mostly a condensation from Guide for School Testing Programs, pg. 25-31.

tions up to the actual date itself. Where norm groups are going to be used as a basis of interpretation, the administration dates should match as closely as possible the time of year the normative data were collected.

Information which is expected to have value for the year's instruction or the student's plans should be collected in the fall. The measurement of progress and the evaluation of instruction should occur in the spring, with the results being forwarded to the school of attendance in the fall.

Scheduling individualized assessment follows the same considerations, except that the individual and the teachers are more directly involved than the group and the school administrators.

C. The Environment

After the test dates have been scheduled, the rooms in which the administration is to take place should be arranged. The room or rooms selected should provide the student with generous working space, freedom from distractions, good lighting, comfortable temperature, and quiet--in short, anything that would allow the student to concentrate comfortably. The rooms should not be disturbed during the testing period. Where the administration is individualized, the settings in which each individual responds to the test should be the same, or very similar.

D. Staff Support

Given the number of tests or assessments of individual or group types administered in a given year and the number of students to be assessed, it is possible to determine how many staff members will have

to devote how much of their time to conducting the administrations. It will be necessary to have central coordinating functions, on site coordinating functions and examiners functions performed. The staff that performs these functions should be well represented on the advisory committee that governs the assessment program.

E. Training Examiners

Accurate test administration technique is a special knowledge and should be systematically taught, learned and relearned each year. The instruction and practice provided should be expertly run and sufficient to give the initiate minimum competence and the experienced refined skills. Video simulation methods are useful.

The content of in-service sessions for administering specific tests or assessment techniques should include detailed knowledge of how to administer the assessment and practice in doing so, covering most of the following points.

(1) Careful study of the test's administration manual.

(2) How to keep accurate time.

(3) The need for slow, clear speech.

(4) The need for rigorous adherence to the standardized procedure.

(5) Preparing the students for the test.

(6) Answering and not answering student questions.

(7) Avoiding disciplinary disruptions.

(8) Cheating.

(9) Reporting events which invalidate the results.

(10) Keeping sufficient supplies on hand.

(11) Keeping track of the testing materials.

(12) Returning the immediate results to the coordinators.

In-servicing sessions of this sort can involve the central coordinator and the site coordinators, as well as the site coordinators and the examiners.

F. Distribution of Materials

The materials required for the testing should be distributed to the site coordinators a week in advance, stored securely and then distributed on the appropriate day. The on-site coordinator should count the materials and record their number on a central form or checklist, also indicating which examiners received and returned how many of each item.

G. Quality Control

Feedback from the examiners on their actions and observations during the actual administration is necessary to assure that the required conditions have been satisfied. This requires a free flow of negative information as well as positive. This requires a trust and rapport among the examiners and coordinators. There may be deficiencies or unforeseen circumstances in the procedures themselves. There will be a need to quickly administer make-up tests. It is very desirable to be advised of events occurring during the administration that would invalidate the results; e.g., misinstructions or untimely disruptions. Since these oftentimes must come as self-confessed errors on the examiner's part, care should be taken to avoid even the slightest gesture of disapprobation. It is essential that negative information be communicated

and not suppressed. The examiners who are cognizant of their own errors and who do not receive disapprobation are those most likely to learn the most from their experience. The student taking the test is the ultimate beneficiary of this rapport.

H. Examiner Checklists

For whatever test or assessment technique, each examiner should have a checklist of items necessary for the proper administering technique. This checklist will vary in item and deed from test to test, but will include information like that in the following illustrations.[2]

1. EXAMINER CHECKLIST BEFORE TESTING

(1) Know the purpose for testing. The person who administers the test, as well as the person(s) who will use the data, should understand the reason for the testing. The person taking the test should also understand the purpose for testing.

(2) Know the test.
   (a) Familiarize yourself with the content and makeup of the test.
   (b) Read the publisher's directions for test administration in order to be completely acquainted with procedures, time restrictions, etc., for each part of the test.
   (c) Know the test and answer sheet in order to respond quickly and surely to questions raised by the person taking the test.

(3) Prepare students for testing--set up a good test climate. Much can be done to assist those to be tested. Well in advance of the test date, the student should know:
   (a) That he will be tested.
   (b) Why he will be tested.
   (c) How it will benefit him to be tested.
   (d) What will be done with test results.

---

[2]The three illustrations, before, at the time of, and after testing, are mostly quoted from Standardized Tests: A Handbook for Administration and Use, Donald E. Lown, Greece Central School District, Rochester, New York, pg. 23, 24, 32, 33.

Students should be comfortable psychologically as well as physically. A certain readiness tension, tone, or set, is good so that the student addresses the test situation with seriousness of purpose. This should not, however, be a frustration or a fearful tension.

Since the purpose behind different standardized tests varies, the mind-set that a student should have in being prepared for a test also varies from one situation to another. It is important to distinguish here two types of tests or test purposes. (1) Standardized tests used within the district: such tests are tools to learn more, both for the individual student's benefit, and for student benefit collectively through program analysis. Students should know that, to serve as a tool, testing offers a benefit rather than poses a threat. (2) Tests with rigid administration requirements such as scholarship qualifying exams (PSAT, RSE, ACT, SAT): toward such tests students should know that individual results carry important reward values. Helping the students to see such tests in the perspective of their own goals should replace threat with purpose.

(4) Prepare testing materials
Have a sufficient quantity of all testing materials (tests, answer sheets, pencils, scratch paper if needed, demonstration materials, and whatever else may be necessary) on hand and well organized. Organization of all materials should be completed at least a day in advance of the testing so that no stress or confusion exists due to management problems as testing begins. A relaxed, self-assured demeanor on the part of the test administrator carries over to the student.

2. EXAMINER CHECKLIST AT THE TIME OF TESTING

(1) Set the proper tone for a good test situation
- (a) Emphasize that students should do their best in answering each question.
- (b) Explain to students about the test (such as, "Some of the questions will be difficult and students are not expected to know all of the correct answers.").
- (c) Let students know some hints for effective test taking, unless expressly forbidden by the manual.
  - (1) Tell students to skip a really difficult question and go on to the next. The object is to show how many they can answer, so they should not get bogged down on one sticky question. Then, if there's time, they can go back and try the ones they skipped.
  - (2) Advise students not to go back and change answers. Their first thoughts are usually best.
  - (3) If guessing is penalized in the scoring, encourage students to answer judiciously. If they think they have any idea of the right answer, fine, but discourage random guessing.

(4) If an "I don't know" choice is available, emphasize that if truly accurate, this is a perfectly proper answer. (It should not be used as a cop-out, however.)

(2) Set up a proper testing environment
- (a) Provide each student with adequate, comfortable seating space. Separate students to prevent distractions and copying.
- (b) Have all desks or tables cleared of materials except for pencils and erasers.
- (c) Post a sign (Testing - Do Not Disturb) to prevent disruptions while testing is underway.

(3) Pass out supplementary materials. Make sure all students have enough pencils, scratch paper, if necessary, etc.

(4) Follow the publisher's directions.
- (a) The directions that accompany standardized tests are very detailed and specific. They should be followed exactly to assure uniformity of testing within the district and for comparability of results to the normative data.
- (b) Explain procedures and directions to students such as:
  - (1) This is a timed test, do not open the test booklet until instructed to do so.
  - (2) There is no passing or failing this test, so do your best.
  - (3) All answers must be put in the appropriate spot on the answer sheet (if separate sheet is used).
  - (4) At this time we will complete only the reading (for example) section of the test.
  - (5) Be careful not to make stray marks on your answer sheet.
  - (6) Here are your test papers, let's do the practice question, etc., etc.
- (c) Ask if there is anyone who does not understand what is to be done.
- (d) Begin the test at a known time; monitor students during the testing; make sure all students stop work on the test exactly in accordance with the specified time limits. Don't by sympathetic and give students an extra minute.

  Time limitations are not punishment techniques; they are to assure everyone an equal opportunity. Time limitation is a way of controlling one of the variables that can affect the ability to compare results in a valid manner.

(5) If more than one section of a test is being administered, allow for a brief rest period between work periods.

3. EXAMINER CHECKLIST AFTER TESTING

(1) Collect and account for all testing materials, keeping test booklets separated from answer sheets.

(2) Scan all answer sheets to assure that each student's name and identification number have been filled in correctly.

(3) Organize materials in the way you want to have the testing results organized. Whether tests are to be hand scored or machine scored, organization saves wasted time, confusion and potential error. If test results are desired by individual class, papers should be grouped this way. If results are desired on a grade level basis for an entire school, they should be so grouped.

(4) Send all answer sheets immediately to the central office so scoring and data processing can begin, excepting hand score materials to be done by the tester. (However, since testing situations may vary from time to time, follow the specific directions relative to each testing situation.)

(5) Prepare yourself and the students for the return of information while scoring and data processing are taking place. What will the results mean to you? How will the information be used?

I. Scoring

There are basically three choices with regard to scoring: Score by hand; send the answer sheets to the publisher's commercial scoring service; or set up a district scoring service.

Hand scoring is frequently preferred by teachers of the primary grades, inasmuch as most of these teachers acquire a very special interest in seeing first hand how their students performed. It is further beneficial since the teachers immediately possess information which may affect their approach to instructing this or that student. The hand-scored tests can be returned as the full test booklet, the front cover only with total raw score, or as a list of names and scores. Conversions of the raw scores onto derived scales such as percentiles should be done centrally where greater clerical efficiency can be achieved once the conversions formulas are set up.

When the scoring is done by a contracted machine scoring service, there will be time lag between the date of administration and the receipt of the results, possibly three to four weeks or longer. Such services, however, are usually cost effective and offer numerous report formats, and sometimes will develop customized formats for individual districts. This should, of course, always be the district's preference. Individual pressure-sensitive labels are available through this option.

The considerations for developing a district or system scoring service are well detailed in Memo No. 5: Operating a Scoring Service, in the NCME's Guide To School Testing Programs. Some of the advantages of such a system, besides scoring, norming and reporting test results, are: the ability to score and analyze teacher-made tests; scoring and analysis of evaluation instruments and opinionnaires; short turnaround; customized and modifiable output; data banking longitudinal data (computer storage); printing or reproducing tests; and compiling unique tests from an item bank. The future reduction of cost and space in computing technology through micro-processors will make it possible for any district to acquire local scoring service capabilities that can be maintained and operated by a licensed electronic technician for an initial cost roughly equal to a single F.T.E. payroll unit.

Types of Scores[3]

1. The Raw Score -- A raw score is a direct numerical report of the individual's performance. It may be the number of items answered

---

[3]Condensed from Essentials of Psychological Testing, Second Edition, Cronbach, 1960, pg. 69-87.

correctly, or a time measure, or some other number taken directly from some direct measure of the performance itself. When the interpretation is to be norm-referenced, the raw scores are converted to percentiles or converted through tables of norms. When criterion-referenced, the interpretation compares the raw score with a criterion or cut-off score, expressed in terms of the raw score's scale.

2. Expectancy Tables -- An expectancy table shows the probable score one may be expected to get on another test, or some other significant criterion of future performance, given this or that raw score of the test administered. For example, a raw score of 120 on a mechanical aptitude test may be correlated with a 70% success rate in automotive maintenance and repair. In diagnostic situations, expectancy tables are often a sufficient rendering of the results. Data for these tables can sometimes be collected locally, e.g., from freshman G.P.A.'s of last year's graduating class. These probability estimates will tend to have a more intuitive interpretation, since the data is collected within a known community for members within it.

3. Percentiles -- A percentile score is the rank of the raw score within a distribution of many raw scores, expressed in percentage terms (see Computing Guide 1). It states what portion of the group represented by the raw scores falls below the score. Percentile conversions made through norm tables compare the individual performance with others outside one's locality. They are especially

useful in diagnosing extreme cases or estimating one's rank in competition with select groups.

4. Standard Scores -- A standard score is computed by transforming the raw score into a Z score. This is done by subtracting the mean of the group from the raw score and dividing the result by the standard deviation of the group. No interpretable information is lost in the transformation. Standard score conversion may result in different scales of different length and spread. The T-score scale with a mean of 50 and a standard deviation of 10 is commonly used (see Computing Guide 3).

5. Normalized Standard Scores -- If one can assume that the norm would produce a normal distribution of raw scores, the standard scale can be normalized by calculating local percentiles and converting those scores into a normalized standard scale through the Table in Computing Guide 4. This conversion will compress a T-score scale into a range of 20 to 80.

6. Criterion Scores -- A criterion score is a point on the raw score scale which signifies the point at which the designer and users of the test decided that one can be reasonably confident that the individual possesses a sufficiency of whatever is being measured. Expectations and standards of mastery are set in criterion scores. Eighty percent is a favorable 4 to 5 odds ratio and a common standard, though mastery standards will tend to be more stringent. (See Computing Guide 5).

7. Error of Measurement -- All measurement has error, and any single point score is apt to include a large quantity of error. The stan-

dard error of measurement can be estimated for any group of raw scores by the following formula.

Standard Error = $S\sqrt{1 - r_{**}}$

Where:

S = The standard deviation of the group of raw scores

r_{**} = The estimate of the reliability of the instrument

S can be obtained locally or is provided by the test manual for the norm group. R should be provided by the test manual, but it also can be computed locally as an internal consistency index.

Given an individual's obtained score, one can be 67% confident that the individual's true score, what he would get if there were no error, falls within a range between the obtained score minus the standard error of measurement and the obtained score plus the standard error of measurement. Do this at twice the standard error and 95% confidence is reached. Since this band will typically cover around 10% of a standard scale, it is almost always desirable to present individual results in bands of probability instead of point scores. Bands can be converted into percentiles by the same procedures used for raw scores, except that two percentiles are produced, indicating the upper and lower boundaries of the band.

8. Profiles of Part Scores -- Most assessment instruments present results in parts--several separately calculated raw scores. The raw part scores will have distinct psychological properties in relation to the other parts, and will behave statistically in a manner of their own. Each part score will be subject to separate reliability and validity estimates, and in comparison with norm groups,

will have its own conversion table. Where profiles are presented graphically, the part scores should all be presented in the same standard scale. Differences between part scores are less reliable than the scores themselves because a difference in scores will contain the two standard errors of measurement, one for each part, instead of the one standard error contained in a single part score.

9. Item Analysis -- By defining groups of students by grades or classrooms, it is possible to generate item analysis, statistics which break down the performance of each individual and the group, item by item. An item list format would present each item and the percentage of the group answering it correctly. An individual-by-item table would present the performance of each individual for each item. When access to automated data processing exists, internal consistency coefficients can be calculated for select groups of items and used as reliability estimates. (See Technical Information, Vol. III, on Reliability.) These estimates may be compared to estimates given in the test manual and used in the computation of the standard error of measurement.

The individual-by-item table is most useful at the classroom level, whereas item list formats are sufficient at the department or school level. The other types of analysis that are useful at the larger group levels are the calculation of group means and standard deviations, usually by grade, class, or school. These averages, when recorded periodically, set up a baseline of expectations and are the most integral part of the accountability function. Such longitudinal data should be reported in many year trends with

an overall average.

J. Record Keeping

The fundamental unit of record is the individual cumulative file which contains a record of all information obtained from measurement, in addition to other information documenting the individual's growth. This file is frequently stored in the principal's office in elementary schools and in the guidance and counseling office at the secondary level. It is typically open to teachers, guidance personnel, and any professional working directly with the student. Given that the file frequently contains personal information of sensitive nature, it is useful to develop two sets of information. One would be an administrative file containing everything except diagnostic information, such as a general ability quotient. The other would be a psychological file containing everything, but open only to restricted persons. Students and parents should have knowledge of this record and right of access to the file. The two-file system does provide some protection of privacy for the individual. Those items of information, e.g., test results, that are placed in the cumulative record, should be defined in written and standing policy, so that one would expect to find the same categories of information in each and every student file.

Automated Record Keeping[4]

There are types of records that can be kept at various group levels:

---

[4]Full treatment of automated data processing can be found in The Guide For School Testing Programs, pg. 32-40.

classroom; department; grade; and school or district. Some of these statistics can be obtained from contracted scoring services and stored in files at the appropriate field unit, e.g., classroom, department, school or district.

The various group summaries can be processed by computer if the district has or has access to computing power and programming assistance. The individual's cumulative file can be reduced to individual records on the computer memory medium (cards, tape, disk and floppy disk). Computerized storage of information is such a convenient commodity that the full record of the individual's locations of attendance and course work can also be stored. With this information as a data base, computer software can be written to analyze and produce summary reports at any specified level of grouping, in any number of report formats. Explanatory text can be commonly stored and called forth for printout whenever required by a report format. Given proper equipment, it is possible to produce the reports in clear, well-spaced, 8" x 11" page layouts.

An organized data base of individual records can also be analyzed internally, producing coefficients of criterion related validity for the tests and techniques as used locally. If the scoring service includes item analysis, the item-by-item statistics of the test over the years can also be stored electronically and summarized at will.

Explanatory Text

To use a file or a report requires that one be able to interpret it and do so correctly. This requires that the user possess information on the elements of interpretation and what it means. It also requires

that the user understand how the results are presented, that is, comprehends instructions on reading the report format. It is desirable that this information be made part of the file or report in such a way as to give the layman a common sense competence and to refresh the expert's skill. One of the major reasons test results are misused, abused or not used at all, is the incompleteness of explanatory information accompanying the results. Such text should exist at least in central files, and if the record keeping is automated, it can be stored and printed out within the context of a report format, where applicable. (See "Elements of Interpretation.")

Long Term Storage

Unless the district bequeaths the individual's cumulative record to the individual upon graduation, the file stays in the district and must either be reduced to microfilm or microfiche, or eventually be destroyed. Other types of reports, item analyses and group summaries may be retained or disposed of at different levels--classroom, department, or school. For instance, handling of individual-by-item reports would be the teacher's responsibility, whereas grade-by-grade summaries would be the school's responsibility. If the summaries fit an 8" x 11" layout, their written forms can be stored in looseleaf notebooks and keyed to the design statement. The counselor can do the same for diagnostic information. The notebooks would, in effect, be the standing historical record of the results of the assessment program at a given point in time for a given population. Mylar cover sheets for each page will increase the durability of the record.

Reporting

The first step in developing a system of reporting assessment information is to identify audiences that will receive, or have access to, the reports. Generally, these will be one of six kinds: the individual; teachers; counselors; parents; administrators; and the public.

Each of the six audiences will have different orientations toward understanding results and will need more or less assistance in reading and interpreting any given result. This first step is what is documented by the specific uses form as part of the expanded chart entry accompanying the Administrations Chart.

The second step, then, is to determine what explanatory information in what form will be necessary for the particular audience to carry out its specific uses in a sound interpretation of the results. General guidelines to this end are:

i. Use non-technical language wherever possible, without underestimating the audience and without being cute.

ii. Use spacious layouts, wherever possible.

iii. Present results in the form of graphs and tables with clear instructions on how to read them.

iv. Report information in the form of confidence bands, based upon the standard error measurement, wherever possible.

v. Be timely; disseminate reports to the audiences before the hour of relevant decisions. Belated information is undesirable.

vi. Involve the potential audiences in planning, wherever possible.

vii. List and generally define each type of report issued.

Common Points of Information

Each report is a whole and, as a document, it should be as self-explanatory as possible. The report should begin with documentation for the test, covering points of information common to all assessment instruments. The results then, the data reported, are presented in one of several types of reporting formats. The common points of information for each report are listed below. Their object is to sufficiently acquaint any interested third party with the test purpose, genre, uses, and relevant population, in order to prevent misinterpretation or misunderstanding of intent.

(1) General purpose

(2) Type of test

(3) Name of test

(4) Groups assessed

(5) Audience of report

(6) Specific uses to which the information should be put

(7) Report format

(8) Type of score

(9) Standard of comparison

(10) Instructions for reading format

(11) Date of issuance

Reporting Formats

(1) The Cumulative Record File -- Individual Reports

When a test result is entered into the student's cumulative record, an individualized report is necessary. Since the entry will

be primarily reviewed by the individual, parents, counselors and teachers, it is very desirable that the report includes all common points of information. This information will insure that the potential interpreter will be apprised of all the elements of interpretation and that all successive interpreters start from the same common basis of information.

The individual report for any one test or assessment will include information on the critical elements of interpretation and the results. This can best be presented in a standard 8" x 11" layout. Insert spaces can be left for affixing pressure sensitive labels. It is desirable to have computer output also constrained to the standard layout. The report then can be stored in an orderly fashion in loose leaf notebooks as well as in files and in folders.

(2) Group Summaries

Whenever the test results are presented as lists of individuals or group averages, the report is a group summary. The groups can be defined in different field units: classrooms; departments; schools; grades; etc. This data will be primarily used by teachers and administrators, but is also the basis of school's public accounting to parents and the community. The common points of information are needed and possibly in greater detail, since the report must communicate with a larger audience. Most group summaries will report the criteria of progress and achievement, but some summarized diagnostic information is of value in instructional and curricular planning.

a. Class Lists -- A class list format is used when the results are

reported for each individual, usually in alphabetical order, in a table crossed by test items, or by the instructional objectives represented by the items, accompanied by row and column summaries. The teacher may be able to use this information to plan with the strengths and weaknesses of the class and its individuals.

b. Trend Tables -- A trend table format crosses a grouping unit, e.g., class, grade, or school, with the averages of similar groups in previous years. In some cases, the group changes every year; in other trend tables, the same group may be followed for several years.

c. Comparison Tables -- An individual score, or a group average, can be compared to either a criterion standard or the performance of other groups. The criterion score may be presented to the right of the score in question, or the reported result may already summarize a comparison made at the time of scoring, e.g., negative - positive, pass - fail, mastery - non-mastery.

Group comparison can be done either by percentile conversions through norm tables or by group averages. In either case, the comparison groups should be tabled to the right of the individual or group score under examination, e.g., individual score, school percentile, district percentile, state percentile, national percentile.

d. Distributions -- The distribution of any group of scores can be portrayed by a vertical numberline that represents levels of performance or achievement on the test. It will denote, at

each level, to percentage of students having scores that fall within the particular levels of achievement. Reference group comparisons can be made with parallel numberlines to the right.

e. Performance Sequences -- Assessment batteries like N-ABELS, or competency measures, represent sequences of critical performances to be demonstrated by the student. These sequences typically exhibit hierarchical characteristics and may be flexibly administered throughout a significantly long time span, e.g., the three intermediate grades of junior high or middle schools. Consequently, at any one time, more or less students will have attempted and passed any particular task in the sequence. The varying time element is easily recorded on the individual record simply by noting the date of administration and the date of passing for each task in the ordered sequence. Group summaries, however, are cumbersome to produce, since they should be done periodically and must be tabulated clerically from each individual record. If the records are kept in computer data bases, periodic group summaries indicating the accumulating percentage of the student body taking and passing a task under measurement in the sequence over the years can be easily produced. Performance sequence data can be reported as class lists, trend tables, and comparison tables.

f. Composite Subject Charts -- A performance sequence or a battery of part scores essentially breaks down the individual's abilities and performance into smaller, but meaningful, units of interpretation. This composite breakdown can be conceived in

other ways, the most common of which would be by specific subject matter achievement, e.g., social studies, mathematics, health, grammar, literature, and so on. Such a profile could include information from several independent tests, although the different types of scores must be clearly distinguished and the exact name of the test footnoted. As a class list, the subject divisions should be the column heading. In group summaries, trend tables, comparison tables, and individual reports, the subject matter breakdowns should comprise the row headings.

ELEMENTS OF INTERPRETATION

There are two objectives in the interpretation of assessment information. The first is to be assured that the information is accurate. (Is the test reliable? Was the administration sound?) The second is to see that the persons interpreting the information are possessed of the explanatory information necessary to form a plausible, common sense, sound interpretation. The groundwork of sound interpretation is first laid at the time of choosing tests, and is reinforced at the time of administration. If the test's qualities well suit the district's purposes and if the administration is well controlled, the results are fit for interpretation. The results become information which may make a difference in upcoming decisions. They do not move from the status of mere numbers, loosely connected to vaguely conceived scales, unless the circumstances surrounding their origin, nature, purpose and use are clear to the reader. This information is the factual grist. Without it, the

reader is left to conjecture. This is the role of the common points of information. They state the general purpose of the information, the nature of the thing measured (e.g., mental quality or behavioral capability), the type of score, the standards of comparison (e.g., norms or criterion scores), and the special uses to which it could be expected to put the information. Instructions on how to read the reporting format (e.g., graphics or tables) are also necessary.

Some general guidelines for interpretation when viewing assessment information are listed below. They begin to apply once the common points of information are understood and the quantitative judgment of more or less has been made.

1. Look for the usual as well as the unusual. Normal or average performances are as useful to future planning as are the extreme cases.
2. Look for trends or patterns. It is useful to discern the individual's pace and breadth of learning as well as to determine recurring strengths and weaknesses.
3. Don't over-emphasize small differences. Keep in mind the error of measurement. Remember that the criterion of significance between part scores and gain scores is twice as strict as the criterion for single scores. (There is twice the potential for error between parts and retests.)
4. Don't expect to be enlightened or to discover something new or different with each student.
5. Consciously balance the judgment with as much other empirical evidence as can be obtained or called to mind. How have other students performed in the past? What do other people say about

the individual's abilities? In what else has the individual excelled, or failed, or performed normally? Is there a need for more information? Any profound diagnosis of an individual's mental ability, or disability, should be based on assessment information drawn from several sources, provided the information is empirically and validly obtained.

6. Make mental notes of interesting findings, or write them down.